General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

NASA CONTRACTOR
REPORT

NASA CR-150221

CMOS ARRAY DESIGN AUTOMATION TECHNIQUES

By T. Lombardi and A. Feller
Advanced Technology Laboratories
Government Systems Division, RCA
Camden, New Jersey 08102

December 1976

Final Report

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Section 1

INTRODUCTION

This report describes the development of a 4096-bit CMOS SOS ROM organized 512 words by 8 bits. A significant feature of this ROM is that it can be programmed either at the metal mask level or by a laser beam after wafer processing has been completed.

Commercially available ROM and PROM chips are made from two technologies: bipolar and MOS. Bipolar ROM chips offer typical access times of from 20 to 150 ns. The 20-to-50 ns range is representative of ECL type designs. ECL chip size is limited due to power considerations, with 1024 bits or less being typical.

The majority of bipolar ROM designs utilize TTL logic. State-of-the-art TTL ROM design offers 16K bits per chip, while PROM design tops out at 8K bits. Access times from 50 to 150 ns are typical, with active power dissipations of 500 to 1000 Mw per chip. Standard TTL voltage requirements (5V ± 5 to 10%) are specified for ROM chips fabricated with this type logic.

ROM chips utilizing MOS technology are made primarily with PMOS, although some NMOS and CMOS chip types are available. As a class, MOS ROMs have access times from about 300 ns to several microseconds. PMOS and NMOS ROMs require from one to three (normally two) voltage supplies to function, while CMOS ROMs require just one supply voltage, which typically can be varied over a wide range (3 to 18 V). Some input and/or output compatability exists between MOS ROMs and TTL circuitry, although most circuit designs utilize special interface circuit elements.

PMOS and NMOS power requirements span the same range as TTL bipolar (500 to 1000 mW), although some chip types have power dissipations in the 150 to 200 mW range. CMOS ROM chips have power requirements that reduce consumption over PMOS, NMOS, and bipolar by a factor of ten . CMOS design sacrifices chip area to accommodate the same number of bits as NMOS, PMOS, or bipolar designs. NMOS ROMs are available with 16K bits and PROMS presently offer 8K bits. The largest CMOS ROM or PROM presently has 1K bits.

The design of modern equipments has moved towards larger degrees of implementation with CMOS circuits. The characteristics of CMOS have been well documented and will nct

be elaborated on here except to point out that the extremely low power dissipation of CMOS has removed a critical hurdle towards the development of LSI circuit types. Improvements in CMOS technology and the maturity of the SOS (silicon on sapphire) process have increased on-chip circuit density and speed to the point where VLSI (over 1000 gates) is possible with system speeds better than low-power Schottky TTL.

One reason for the polularity of CMOS is that it has a wide operating voltage range (3 to 15V) that can be varied by the system designer to accomodate power, speed, and interfacing problems. High-speed CMOS-SOS LSI systems are typically operated at voltages above 5 V in order to realize an increase in circuit speed. When such systems required RAM and ROM components previously, the only speed compatible memories were bipolar. Recently, RCA has introduced a series of CMOS SOS RAM chips that provide speed compatibility with bipolar RAMs, while having the low power and wider operating voltage characteristics of CMOS. There is, however, no comparable (P)ROM component on the market. A 10 V CMOS SOS LSI system requiring a high-speed PROM or ROM is forced to use a bipolar product. This has at least three negative effects: (1) an extra power supply voltage (5V) is required; (2) the low-power CMOS SOS system has been compromised by the use of higher powered TTL; and (3) components in addition to the ROMs are required to interface the input and output voltage levels of the bipolar and CMOS subsystems.

An alternative to using bipolar ROM chips in high-speed, low-power systems is to develop a ROM using a high-speed, low-power technology. CMOS SOS is such a technology.

The development of a 4K CMOS SOS ROM fills a void left by available ROM chip types, and also makes the design of a totally CMOS major high-speed system more realizable.

Section 2

DESIGN OBJECTIVES

A. ACCESS TIME

The 4K CMOS SOS ROM designed by RCA was designated the ATL078. Its organization is 512 words long by 8 bits wide

The design philosophy behind the development of the ATL078 was to make it speed compatible with existing bipolar ROMs at the system level. Specifically, this was interpreted to mean that the CMOS ROM operating at +10 V should provide access times comparable to a bipolar ROM (operating at 5 V) interfacing with a +10 V CMOS system. A nominal time allotted for a bipolar system access was 180 ns. This was broken up as: 10 ns, 10 V to 5 V conversion; 30 ns, address buffering; 120 ns, worst-case ROM access; and 20 ns, 5 V to 10 V conversion. A comparable CMOS ROM system access would avoid the two-level shifting stages so that the system access would break down as: 30 ns, address buffering and 150 ns, CMOS ROM access.

The origin of the bipolar system access values came from a bipolar MROM* memory in the SUMC-IIIC computer. The SUMC-IIIC is a CMOS SOS LSI computer whose contribution is primarily CMOS except for bipolar ROM memories, main memories, and associated interfaces. The bipolar PROM being used in the SUMC-IIIC was the Intel 3604L-6, which has a worst-case access time of 120 ns over a temperature range of 0 to 75°C. The target access time of the ATL078 was determined to be 150 ns as shown previously. If this were to be a worst-case access at 75°C, then derating the CMOS ROM at 0.3% per °C yielded a target worst-case access time for the CMOS ROM of 130 ns or better at 25°C. This was expected to yield typical access times of 50 to 70 ns. The design philosophy was to make this ROM purely static in nature so that the cycle time and access time had target values that were identical.

B. PINOUT

The pinout selected for the ATL078 was influenced by the pinouts of existing 4K bipolar ROM chips. The pinout chosen was that of the Intel 3604L PROM (which is identical with the Intel 3304L-6 ROM). This would allow the SUMC-IIIC to be used as a test bed for samples of processed ROM chips. This pinout also came within one pin of being directly compatible with the MMI 5340, the Intersil 5605, and the Harris 7643-5.

---

*Microprogram ROM

The pinout selected is shown in Table 1. Of the 24 pins on the package, 23 were used. The pin requirements for this chip were: 9 address lines, 8 output lines, 4 chip select lines (2 positive and 2 negative), 1 ground line, and 1 voltage line. The 9 address lines selected 1 of 512 words, each word being 8 bits wide. The 4 chip select lines were required to maintain compatability with the 3604L-6. Of the 4 lines, 2 were positive logic and 2 were negative logic. All 4 lines were internally decoded to control the 8 output drivers. This provided for a fast access from ship select that was estimated to be 30 ns. The voltage and ground pins provided inputs for the CMOS voltage levels of from 3 to 15 volts. It was considered desirable that the chip be able to work over the full CMOS voltage range to enhance its system potential.

TABLE 1. ATL078 PINOUT

Pin	Function	Pin	Function
1	A_7	24	N/U
2	A_6	23	A_8 (MSB)
3	A_5	22	V_{DD}
4	A_4	21	$\overline{CS}_1$
5	A_3	20	$\overline{CS}_2$
6	A_2	19	CS_3
7	A_1	18	CS_4
8	A_0 (LSB)	17	O_8 (MSB)
9	O_1 (LSB)	16	O_7
10	O_2	15	O_6
11	O_3	14	O_5
12	GND	13	O_4

C. OUTPUTS

Two output types are available from bipolar TTL ROM chips: open collector and tristate. In a CMOS design, tristate outputs, which seemed to be a clearly superior choice, can be designed in CMOS with little cost in terms of chip area. The benefit

offered by tristates is a reduction in system power due to the elimination of output pull-up resistors. Since reducing system power is consistent with a basic design objective in going to a CMOS structure, tristates were used for each of the 8 outputs of the ATL078.

The one precaution required of a tristate design was to reduce current spiking to a tolerable and nondestructive level. Current spiking occurs when two or more tristates turn on simultaneously, pulling to opposite logic levels. This happens frequently during the transition between chip select (CS) states. The higher On resistance of MOS devices (compared to bipolar) typically reduces this spiking current to tolerable levels for all but extremely large output transistor sizes. Without placing undue restrictions on the sizes of the output tristate devices, it was felt that a CMOS tristate output could be designed to drive typical system loads (40 to 80 pf) while still maintaining the target system speed.

D. PROGRAMMING OPTIONS

A basic goal in the design of this 4K CMOS ROM was to make it programmable at the metal level either by modifying the metal mask or by using a laser beam to cut metal links on a completely processed chip.

Programming on a ROM by fabricating unique metal masks is common practice employed by many ROM manufacturers. In the CMOS-SOS process, the metal mask is the next-to-last mask, followed only by the mask that opens holes in the passivation layer. This means that a large portion of the chip processing has been completed before the metal programming mask is required to define a unique chip. This could be used as a mechanism to allow for partial processing of a number of wafers before specific ROM types were required, thus providing for faster turnaround after a unique metal mask has been defined. The processing time required for the final two mask types would define the delivery time for chips programmed this way.

As an alternative to programming with unique metal masks, a conventionally processed SOS ROM could be programmed using a directed laser beam. The SOS technology is ideally suited to such an approach since the epitaxial silicon islands that form transistors are normally separated from one another by the sapphire substrate. Sapphire is a very hard, transparent material that acts as a surface on which to grow and a dielectric isolation between adjacent transistors. It also acts as a fine surface on which to cut metal lines with a laser, since no damage will be done to the active transistor semiconductor. Thus, the chip performance should not be affected by the programming operation if sufficient area is left on the sapphire to sever the programming links.

Test were performed using a xenon laser on CMOS SOS 4007 equivalents (dual complementary pair plus inverter). The laser had a 0.2-mil kerf. Numerous cuts were made through the metal interconnect (12,000Å) on several chips. Subsequent probing of these chips indicated no degradation of transistor performance compared to the pre-cutting measurements.

An automated laser programming capability would include a laser source and a programmable transport system. After the initial expense of setting up such a system, the programming operation would be comparable in speed and efficiency to programming stations now available for MOS and bipolar PROM chips. Aside from fast turnaround, small quantity runs of numerous types would become feasible and practical since the ATL078 could be treated as either a PROM or a ROM.

If the ATL078 were treated as a PROM, then the entire wafer processing operation could be completed long before the chips were to be programmed. To operate in this manner, some provision should be made to allow for a pretesting capability on the unprogrammed chip. Some possibilities for pretesting were: (1) power On and monitor leakage current; (2) addition of extra words of memory to allow for a partial test of the address decoders and output drive circuitry; or (3) some variation or combination of the previous two possibilities that would provide as much information about the functional operation as was reasonable without seriously impacting on the chip design. Before the chip design was undertaken, it was recognized that some provision for pretesting should be made; however, it was decided to focus on arriving at an optimum chip configuration before pretest options were added. At that time, a tradeoff could be made that would allow implementation of the pretest options that would have the least impact on the chip area.

E. POWER DISSIPATION

The power dissipation of CMOS is composed of a static and a dynamic component. In normal CMOS design, the static component consists of the sum of the leakages through the Off transistors in each of the complementary structures throughout a chip. The dynamic component is equal to the sum of the CV^2f loses over the chip. The only chip-constant term in the CV^2f expression is the capacitance (C), which represents the gate and the interconnect overlap capacitance. The operating voltage (V) and the operating frequency (f) are user dependent, making the dynamic dissipation of CMOS parts strongly a function of the system in which they are used.

At the onset of this chip development, the chip circuitry was assumed to be totally CMOS except for the 4096 memory elements. The memory elements were conceptualized as NMOS transistors whose source was tied to either the V_{DD} (highest chip potential) or V_{SS} (lowest chip potential). Accessing an MNOS memory element would bring its drain either to V_{SS} or to within one threshold drop of V_{DD} (source follower). In accessing a logic "1" state from an NMOS transistor in this fashion, its drive capability would be reduced and the delay associated with this element would be increased. An alternative to this approach would be to assist the NMOS device in pulling up to a logic "1" state by using a biased-on PMOS device whose source was tied to V_{DD}. This would provide for a faster access, but would also increase the dc chip power requirement.

The overriding philosophy guiding the development of this chip was that it must be speed compatible at 10 V to a comparable bipolar chip (operating at 5 V). Since this chip

was to be CMOS SOS, a power saving of better than 10 to 1 could be expected over comparable bipolar parts. If, however, it were necessary to sacrifice some power in order to maintain the speed objective, this would be done, since the resultant part would still be considerably lower powered than a bipolar ROM. Since bipolar ROM chips require from 500 to 1000 mW to operate, a 10-to-1 power saving would still enable a CMOS SOS ROM to dissipate 50 to 100 mW. This range of value was taken as a design goal and was meant to include both static power dissipation and dynamic power dissipation.

The static leakage of a CMOS SOS chip having 5000 to 10,000 transistors should be 100 to 500 μa at 10 V. This corresponds to a 1 to 5 mW static dissipation. For a CMOS ROM design having no pull-up devices aiding the NMOS memory elements, this represented a realistic goal. Adding pull-up devices to the NMOS memory elements would increase the static dissipation; however it was expected that no more than 30 mW should be required. This left 20 to 70 mW as a target value for dynamic dissipation.

F.   IMPLEMENTATION

The circuit implementation of this chip was to be totally silicon gate SOS. Design rules used in laying out the chip were to be the standard CMOS SOS rules. The validity and maturity of these rules has been proven by close to one hundred successfully fabricated chip types.

The implementation of all input, output, decoding, and buffering circuitry was to be done in CMOS. The only section of the ROM that would not be complementary MOS would be the memory elements. These elements were to be NMOS, either implemented as a low-power array or aided by pull-up devices.

Final decisions as to the size of all transistor stages and as to the method of interconnection were to rely on computer simulation programs. Of particular importance in this area was the speed-power tradeoff associated with outputting logic "1" data from NMOS memory array.

Section 3

CIRCUIT DESIGN

A. GENERAL

The memory array of the ATL078 consists of 4096 NMOS transistors whose sources are connected to either V_{DD} or ground. Circuit implementation of this array is achieved by separating the 4096 transistors into 8 separate blocks of memory. Individual blocks consist of 64 words by 8 bits (see Fig. 1).

Each 64-by-8 block is driven by its own 1-of-64 decoder, whose inputs are the least significant 6 bits of the memory address, A_0-A_5.

The 3 most significant address bits, A_6-A_8, are used as the inputs to a 1-of-8 decoder. The outputs of this decoder control the multiplexed output lines of the eight 64-by-8 memory arrays.

Eight bus lines connect the outputs of the 64-by-8 memory arrays. Each bus line feeds sensing and shaping circuitry, whose output drives a tristate buffer.

The tristate buffers are controlled by the decoder of the 4 chip-select (CS) lines. The chip-select lines are decoded on-chip as an AND function. Using these 4 chip-select inputs, up to 16 chips (8K words) can be stacked before external decode circuitry is required.

B. NMOS MEMORY ARRAY

The NMOS memory array is broken up into 8 64-by-8 slices. Each slice has 64 rows and 8 columns of NMOS transistors. Every transistor in the array has its source connected to either V_{DD} or ground in the programmed state.

Figure 2 is a representation of the 64-by-8 NMOS memory array interconnect. Metal lines run vertically, bringing V_{DD} and ground to each NMOS element. Additional vertical metal lines carry the output signals from the memory elements to the output multiplexer. Row-select lines run horizontally in the array. The row-select lines are polysilicon lines, which serve as the transistor gates as well as the second level of interconnect in the SOS process.
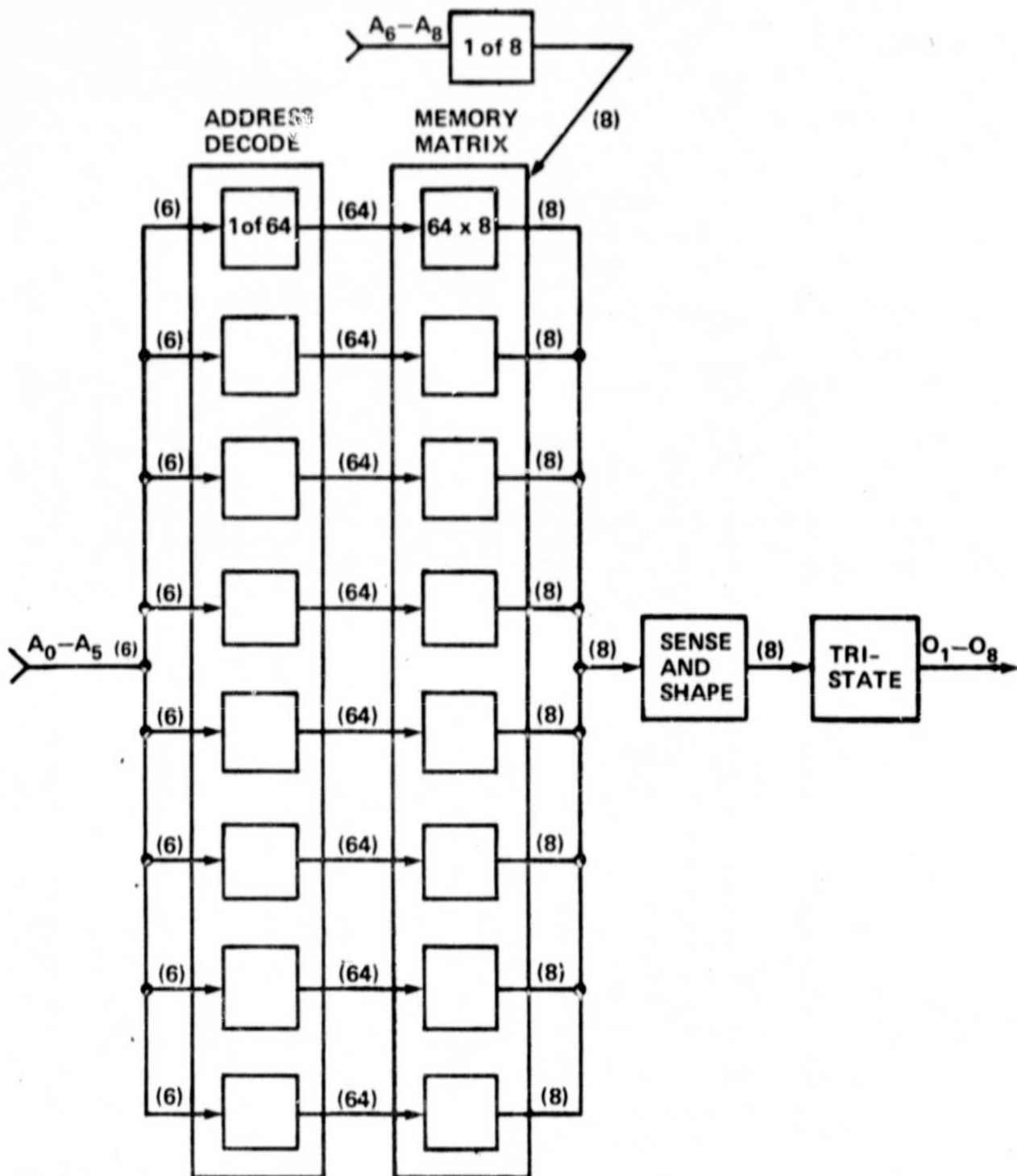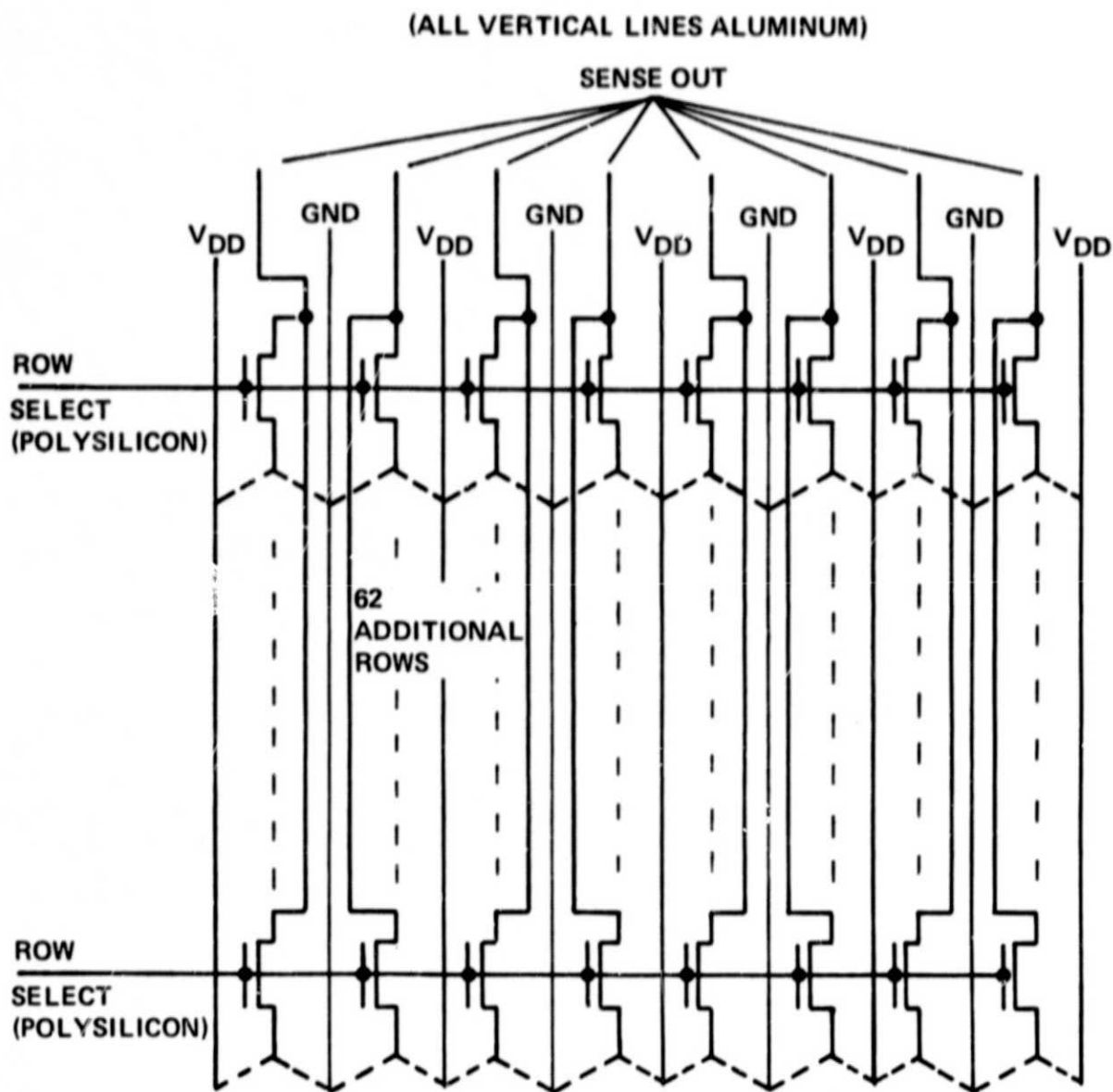
Fig. 1. ATL078 block diagram.

Fig. 2. NMOS array interconnect.

Each row-select line acts as the common gate for 8 NMOS transistors. In the 64-by-8 slice, 64 row-select lines are required. The drains of each of the 64 NMOS devices in a column are connected together by a metal sense line. Eight sense lines run vertically in each 64-by-8 slice.

When the ATL-078 is addressed, one row in each of the 64-by-8 memory slices is selected, while the remaining 63 rows are unselected. Each of the 8 bits on a selected row places its data on a sense output line. Only one NMOS drain of the 64 drains connected to a sense output line controls the state of the line at a time.

Figure 3 shows a section of the NMOS memory array. Eight complete memory cells are shown (in a 4-by-2 array). The cell size is 2.5 by 1.55 mils. Each NMOS device in the memory array has a gate width of 1.6 mils and a gate length of 0.25 mil.

The 8 cells shown in Fig. 3 each have their two programming links intact. This would represent the case where laser programming was to be performed at a later time.

Figure 4 is a "blowup" of the programming links for 4 cells, each of which shows a different programming state. The upper lefthand cell has both links intact, while the two righthand cells each represent cells that were programmed at the metal mask level (one tied to GND and one tied to +V). The cell in the lower left of Fig. 4 is representative of a link that has been laser programmed. The vertical V_{DD} and ground lines buses are each 0.1 mil wide, as are the programming links. Separation between the voltage or ground line and the NMOS epitaxial silicon is 0.2 mil. This 0.2 mil is the programming area for a laser where metal exists over sapphire. A registration and alignment error of 0.1 mil is allowed in any of the four axes. If the 0.1-mil error occurs in the direction of the epitaxial silicon, then it is possible that the laser may cut some metal over the epitaxial material. The epitaxial silicon involved in this area is not in the conduction path between the NMOS source and NMOS gate, so that circuit performance should not be affected.

The design chosen for this ROM utilizes 8 64-by-8 memory slices and 8 1-of-64 decoders (one decoder driving each memory slice). The design could have been implemented using 4 decoders if each memory array were 64-by-16; taken further, only one decoder would be required if the memory array were 64-by-64. The deciding factor as to the final implementation was the access-time requirement of the chip. Referring to Fig. 3, it can be seen that the row select lines are polysilicon.

In a 64-by-8 slice, each row-select line acts as a series gate for 8 NMOS transistors. The resistivity of N type polysilicon is taken as approximately 200 ohms per square. This series resistance, combined with the NMOS gate capacitance, creates an RC time constant whose worst-case delay for a 64-by-8 slice is 12 ns. Increasing the memory slice by 2 from 64-by-8 to 64-by-16 increases the RC delay down the row-select line from 12 ns to 48 ns. While it was felt that 12 ns could be tolerated for this

SENSE OUTPUT SENSE OUTPUT V_{DD} PROGRAMMING LINKS FOR 1 BIT GND V_{DD} SENSE OUTPUT SENSE OUTPUT

ROW SELECT LINES (POLYSILICON)

CELL SIZE
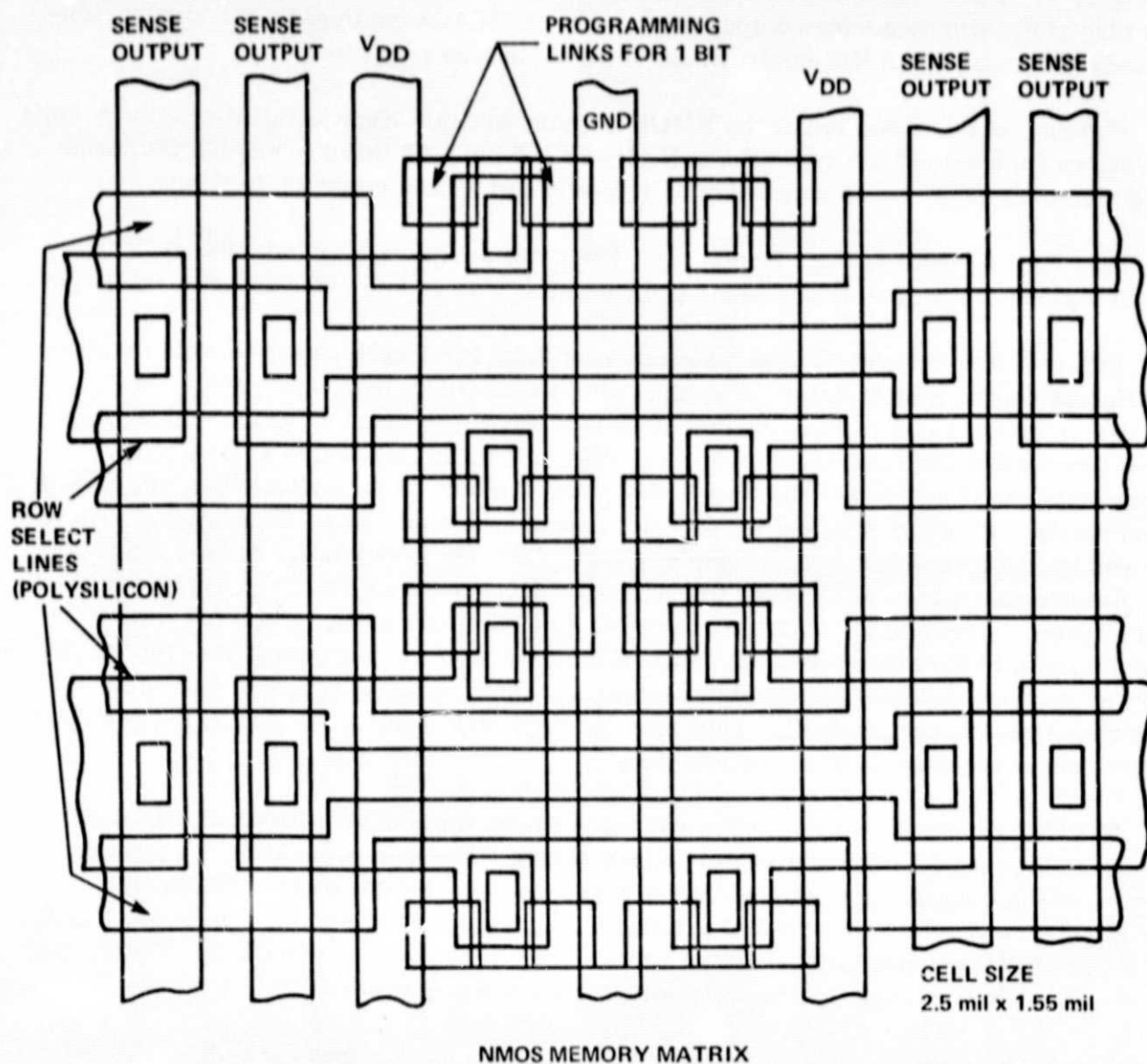2.5 mil x 1.55 mil

NMOS MEMORY MATRIX
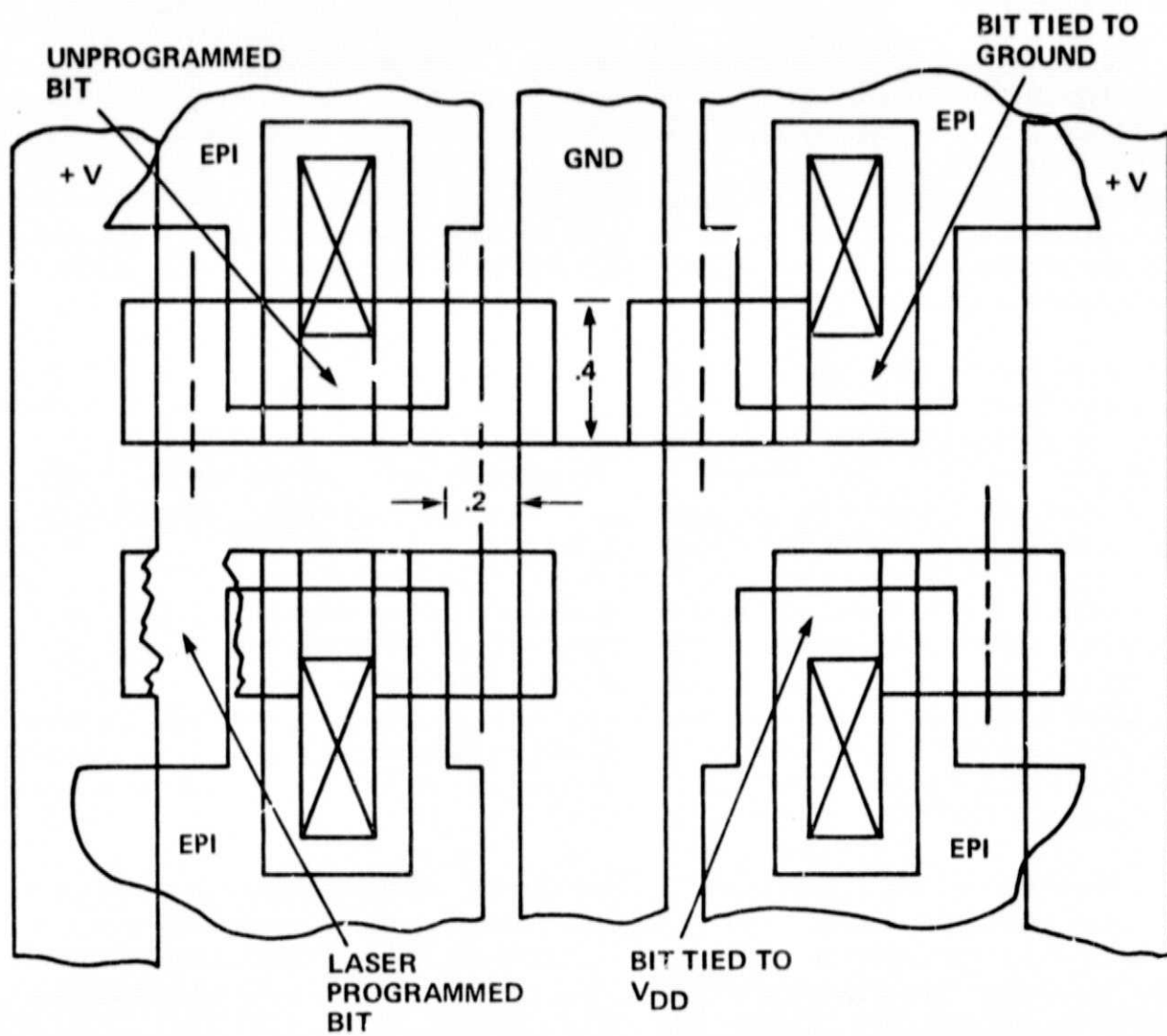
Fig. 3. NMOS memory matrix.

12

Fig. 4. Memory programming links.

portion of the access, 48 ns would have slowed down the overall access so that it would have been over the target goal of 130 ns.

C. 1-OF-64 DECODER

Eight 1-of-64 decoders were used to drive the 8 64-by-8 NMOS memory slices of the 4K memory. The 1-of-64 decoder was functionally implemented as 64 6-input NOR gates with 6 series PMOS and 6 parallel NMOS transistors forming the NOR function. Because of the regularity associated with this type of deocder, the series PMOS portion was able to be implemented as a large tree type decoder.

A section of the PMOS portion of the 1-of-64 decoder is shown in Fig. 5. Ten series transistors are defined by the 10 vertical polysilicon gates shown in the figure. The 8 rightmost gates are address bits A_2-A_5 and their complements. Since each address bit utilizes either a true or a complement state for each of the 64 decode conditions, cnly 4 of the 8 gates can be used for any one address decoder. The remaining 4 gates are eliminated from a particular decoder by short-circuiting the source to the drain of the transistor formed by the excess gate. The short-circuiting links for 4 of the decode conditions are shown in Fig. 5.

The polysilicon forming each of the 8 gates, A_2-A_5 and $\overline{A}_2$-$\overline{A}_5$, extends the entire height of the decoder array, which is 99.2 mils. Two gates are required for each address line, one for the PMOS section and one for the NMOS section of the decoder. Each gate is contacted twice along its width to reduce the RC time constant associated with distributing the address state along the entire polysilicon gate width. One such contact to the A_2 address gate is shown in Fig. 5.

Address A_0, A_1 and their complements occupy only two polysilicon columns in the PMOS section of the decoder. The leftmost polysilicon gate in Fig. 5 is A_0. It extends for 48.9 mils or approximately half of the height of the decoder. Address $\overline{A}_0$ also extends for 48.9 mils but occupies the same column as A_0, covering the other half of the decoder column. Both A_0 and $\overline{A}_0$ are contacted twice. The A_1, $\overline{A}_1$ column alternates as $A_1\overline{A}_1A_1\overline{A}_1$. Each of the 4 transistor gates is 24.1 mils wide and is contacted once.

The NMOS portion of the decoder has ten vertical polysilicon gates in a scheme similar to that of the PMOS section. Six parallel NMOS devices are formed by defining either the presence or absence of epitaxial silicon underneath complementary gate addresses. The NMOS and PMOS sections of the 1-of-64 decoder taken together are 23.1 mils wide and 99.2 mils high.

The 6 series PMOS transistors forming each portion of a 6-input NOR gate for the 1-of-64 decoder have a net effective transistor width of 0.57 mils. This results from 6 series PMOS transistors having the following widths: 48.9, 241., 11.7, 5.5,
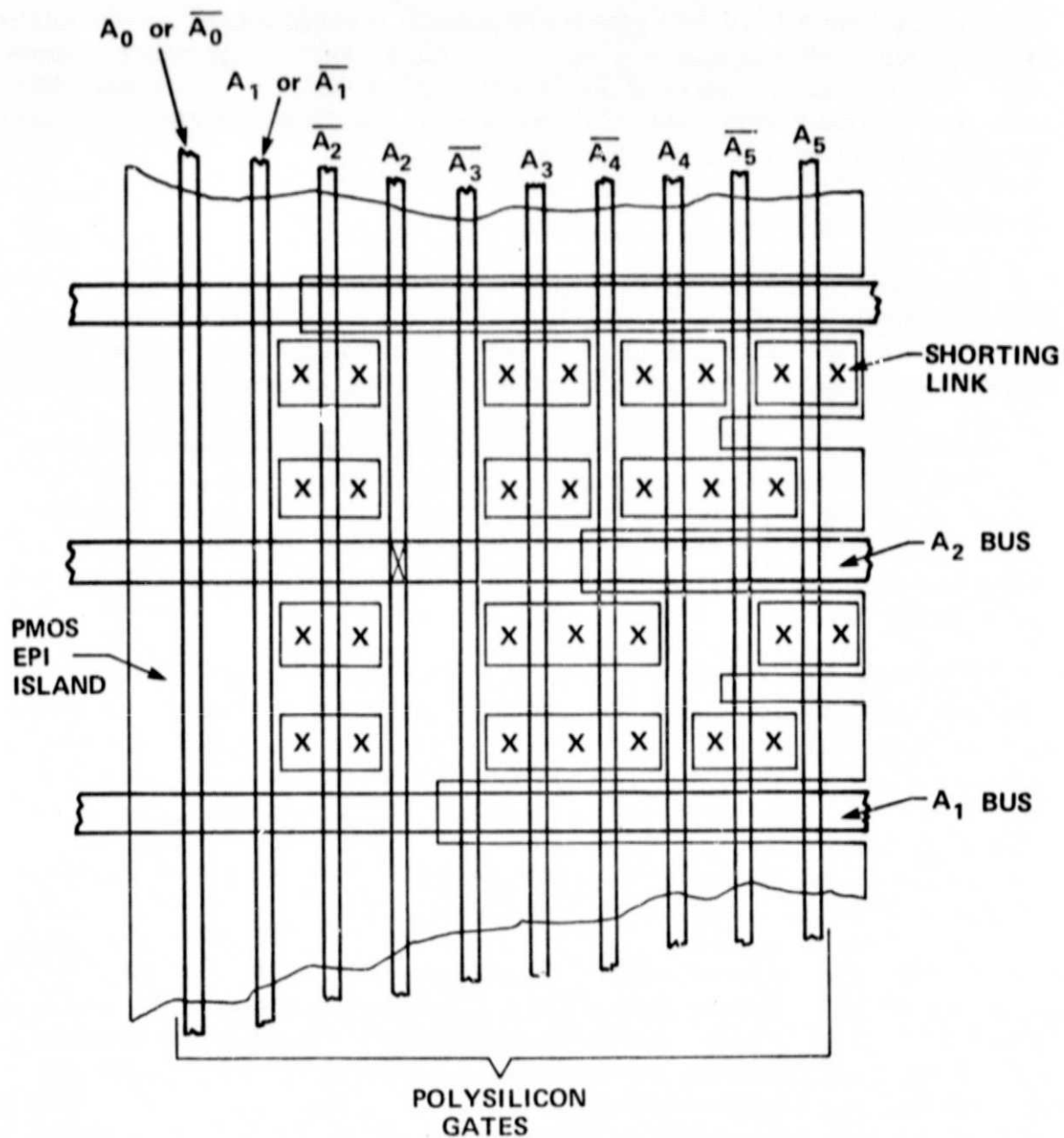
Fig. 5. Section of PMOS structure of 1-of-64 decoder.

2.4 and 1 mils. Each section of the NMOS portion of the decoder consists of 6 parallel NMOS transistors. Each NMOS device is 0.8 mil wide.

The resultant 6-input NOR gate is unbalanced, having over twice the capability to drive a logic "0" compared to a logic "1". Since a logic "1" selects a memory row for interrogation, a previous address should be unselected before new address select information is generated. This will save power, since two NMOS devices in the same column should not simultaneously be On.

D. OUTPUT DECODE

Each 64-by-8 NMOS memory slice is separated from the output bus by a group of 8 transmission gates (one for each bit). Data bus information may come from any of the 8 64-by-8 slices. Only one group of transmission gates will be turned on at a time so that only one 64-by-8 memory slice will control the output data bus.

Command controls for the 8 groups of (8) transmission gates come from a 1-of-8 decoder. This decoder acts on the three most significant address bits, A6-A8, to provide an enable signal for one of the 8 banks of transmission gates. The other 7 banks of transmission gates are turned off so that their associated memory bits are isolated from the data bus. The logic for the 1-of-8 decoder is shown in Fig. 6.

Eight output data bus lines connect each of the common outputs of the 8 multi-plexed 64-by-8 memory slices. Computer simulations at 10 V were made from a row-select input of the NMOS memory slice to the output data bus lines. Other simulations of the critical data path had revealed that a 35-ns access was required for this section if the desired chip access time was to be achieved. The results of the simulation for this section of the data path indicated that a worst-case access of 60 ns could be expected accessing a logic "1" from the NMOS memory elements. Accessing a logic "0" took only 20 ns.

In order to speed up this section of the access time, 8 PMOS transistors were placed on the output data bus lines. The PMOS devices had their drains tied to the output data bus lines, their sources tied to V_{DD}, and their gates tied to ground. Several transistor widths were tried; however a width of 0.5 mil (L=0.25) performed the best. With these pull-up devices inserted into the simulation, the access time to a logic "1" or logic "0" was 35 ns. In addition, the logic swing on the output data bus lines was the full supply voltage, thus increasing noise immunity over the case where no pull-up devices were used. This is a more important consideration if operation at the lower end (3 to 7 V) of the operating voltage scale is anticipated.
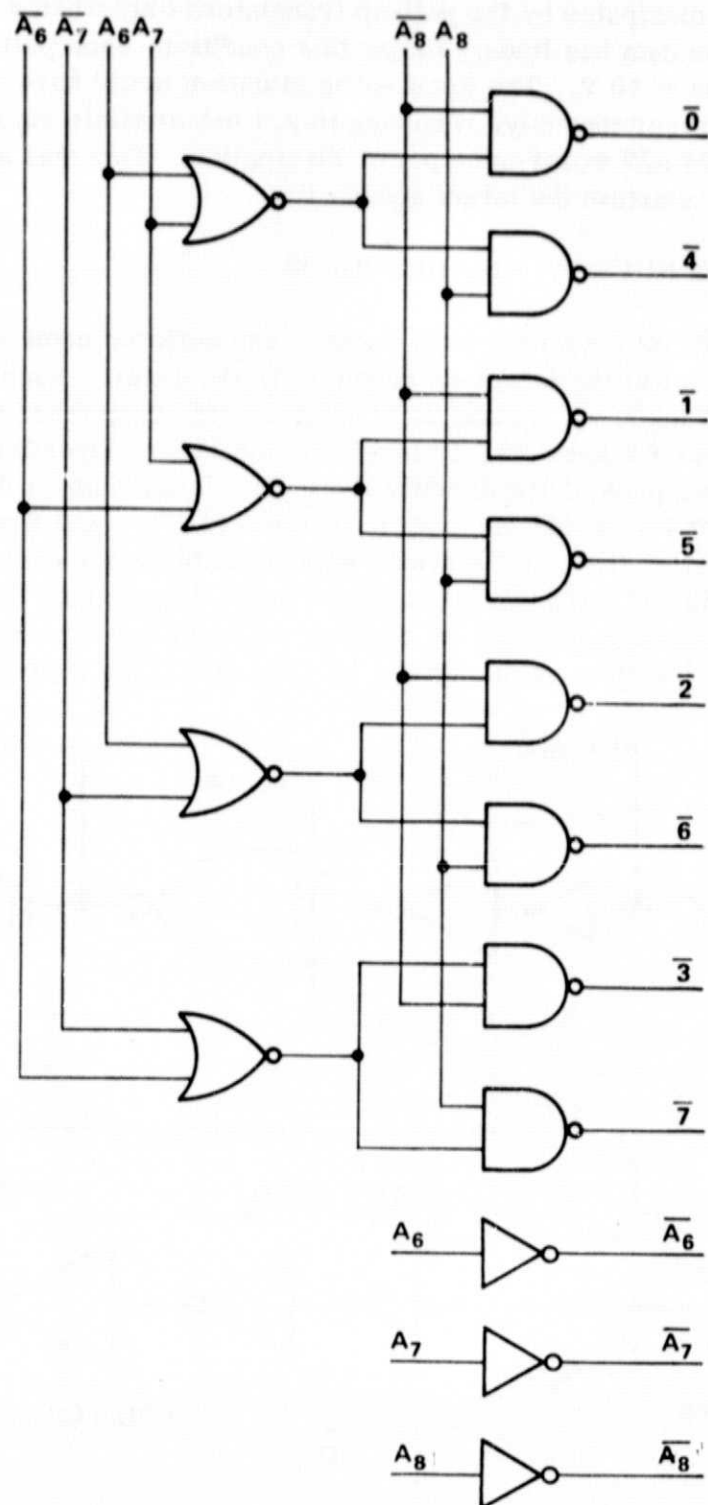
16

Fig. 6. 1-of-8 decoder logic.

17

Static power is dissipated by the pull-up transistors only when a logic "0" is being outputted to the data bus lines. Under this condition, each pull-up device will require 0.3 mA of dc at 10 V. The worst-case situation would have all 8 PMOS devices conducting simultaneously, resulting in 2.4 mA of static current. At 10 V, this represents 24 mW worst-case power dissipation. This was an acceptable tradeoff in order to maintain the target access time.

E. INPUT/OUTPUT BUFFERS AND DECODERS

The least significant 6 address bits, A_0-A_5, are buffered upon coming on-chip and are then fanned out to the 8 1-of-64 decoders in the array. Each 1-of-64 decoder sees 24 inputs: the 6 address bits and their complements each taken twice. At the input to each of the 1-of-64 decoders, 12 inverting and 12 non-inverting buffers shape the address information and drive directly their associated 1-of-64 decoders. Figure 7 shows the data path for an address access on the ATL078. The first address input stage is a large inverter that must drive 32 stages (4 stages for each of the 8 1-of-64 arrays). The second and third inverters shown in the data path of Fig. 7 represent a non-inverting input stage to the 1-of-64 decoder. Two of this type stage and two single inverter stages are driven by each address line on each of the 1-of-64 decoders.
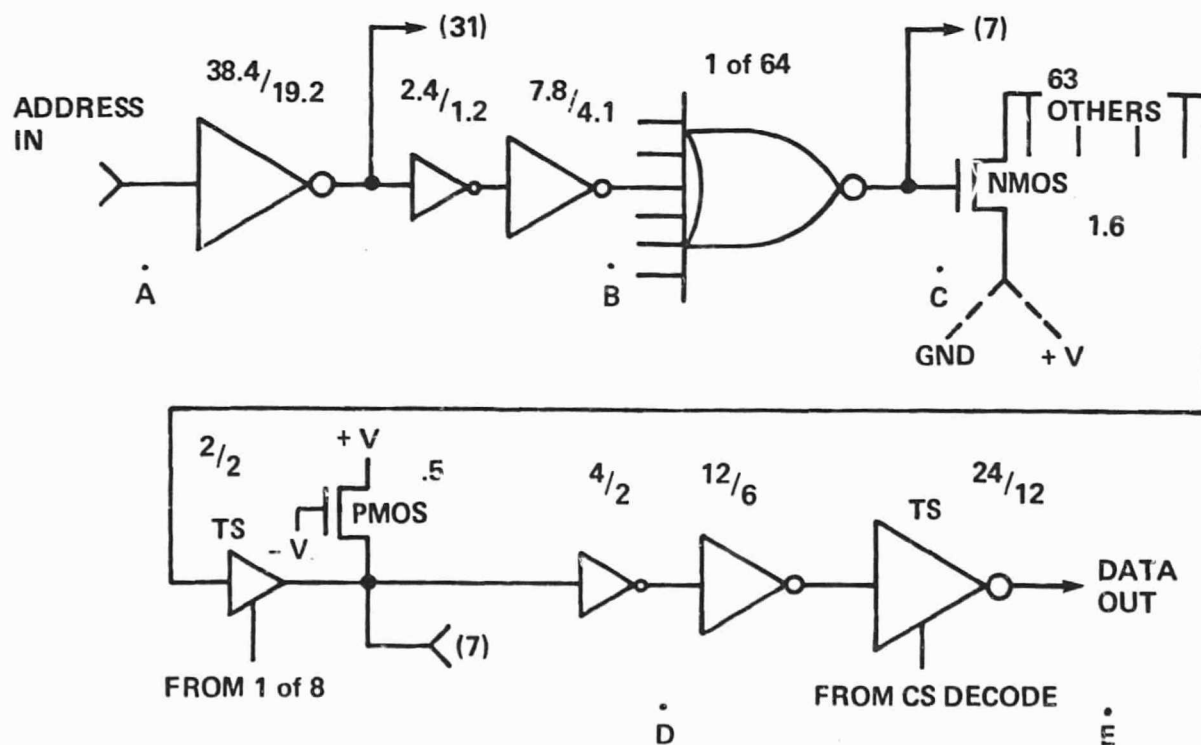

Fig. 7. ATL078 data path.

The output stages of the ATL078 are also shown in Fig. 7. Each data output bus that connects the multiplexed outputs of the 64-by-8 slices drives a balanced inverter, which acts as a sense amplifier. The rise and fall time of the input to this inverter is slow, so the sense amplifier is kept small to reduce loading. Following the sense amplifier is an intermediate buffer that further shapes the output data waveform and builds up the drive capability to drive the output tristate. The output tristate logic and the chip-select decode circuitry is shown in Fig. 8. When the tristate turns on, data enters the tristate logic and passes through input transmission gates to an inverter. When the tristate is in its high-impedance state, the input transmission gates turn off, thus isolating the memory data from the output inverter. Inputs to the NMOS and PMOS elements of this inverter are held low and high respectively by single-ended MOS devices.

The chip-select decode circuitry is a buffered NANDing of the four chip-select inputs. The output of this circuitry is active low and drives all 8 tristate output buffers. Each tristate stage has its own converter to form the required complement of the chip-select enable signal.

All chip inputs and outputs are protected from static charge by diode stacks, which are connected from the signal to either V_{DD} or ground. The diode stacks consist of 4 sets of back-to-back diodes, having a typical voltage breakdown of 24 to 30 V. Each diode in the stack is 2 mils wide.
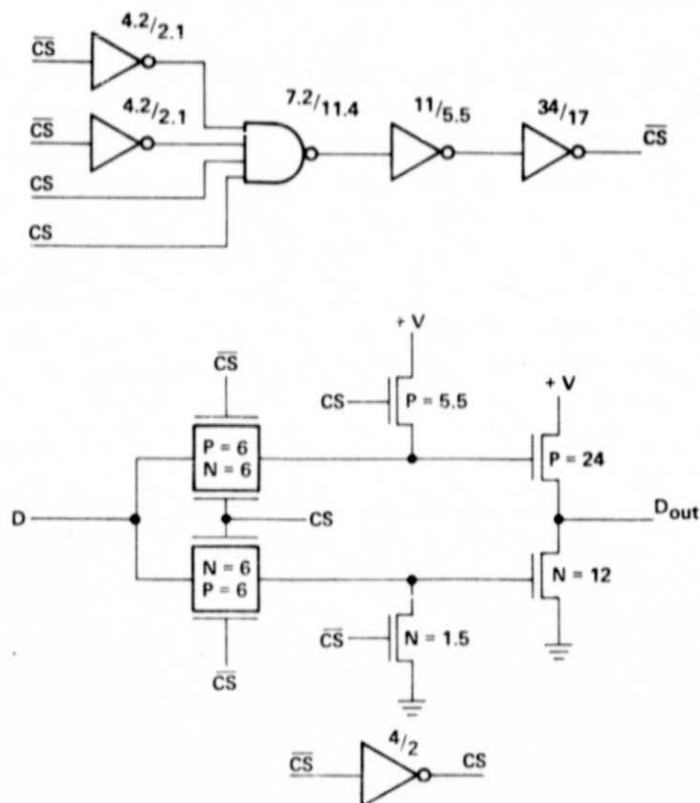

Fig. 8. Chip select decode and tristate logic.

In addition to the input–output diode stacks, a single diode stack is placed between V_{DD} and ground. This stack also consists of 4 sets of back-to-back diodes with each diode having a width of 10 mils.

F. LAYOUT

The block layout of the ATL078 is shown in Fig. 9. The step-and-repeat size is 252 by 257 mils.

The periphery of the chip contains the input inverting buffers for address bits A_0 through A_5, the 1-of-8 decoder using address bits A_6 through A_8, the chip select decode circuitry and the 8 output tristates.

The center of the chip can be considered as 4 quadrants, each containing 1K of memory. In actual implementation, each quadrant contains 2 64-by-8 memory slices, 2 1-of-64 decoders with associated input (inverting and non-inverting) buffers, 2 banks of data-out multiplexers (transmission gates), 2 sense amplifiers, and 2 intermediate buffers.

Figure 9 shows the 2 64-by-8 memory slices as one 64-by-16 block. The 2 slices are connected together, sharing common polysilicon row-select lines. This can be done without increasing delay, since each row-select line is driven from both of its ends by 1-of-64 decoders. What results is an averaging effect that reduces the delay to the centermost (worst-case) NMOS bits of the 64-by-16 memory slice.

Interconnect and I/O wiring on the chip occupies the channels surrounding the perimeters of the 4 quadrant sections. Address inputs to the 8 1-of-64 decoders are routed in the 3 vertical channels on the chip, as well as the topmost horizontal channel. Output control information from the 1-of-8 multiplexer controller also runs in these channels. The 8 data buses connecting the multiplexed outputs of the memory slices run in the center horizontal channel. The sense amplifiers and intermediate buffers for each of the 8 data lines are in the blocks labeled OUTPUT DECODE in Fig. 9. Output data from the intermediate buffers to the tristate drivers is routed down the 3 vertical wiring channels and across the bottom horizontal channel.

G. TESTING

A pre-programming test capability was incorporated into the design of the ATL078 for chips that are processed to be laser programmed. Chips that are programmed at the mask level require no special test circuitry and may be treated as normal memory chips for test purposes.
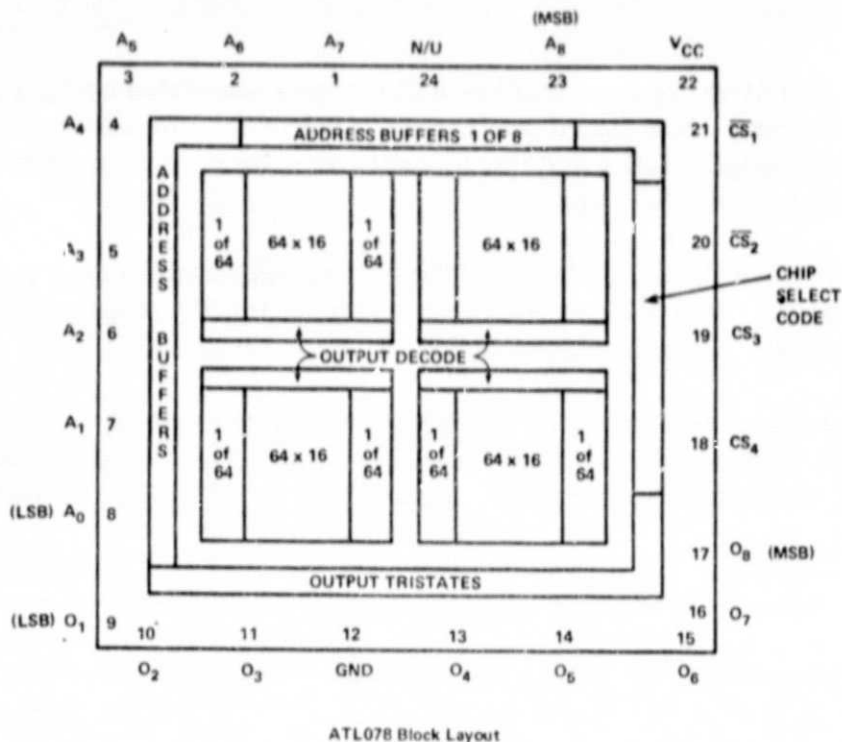
ATL078 Block Layout

Fig. 9. ATL078 block layout.

Chips processed for laser programming have two metal links tied to the source of each memory bit, one link tied to ground and one link tied to V_{DD}. As a result, a direct short circuit between V_{DD} and ground exists, preventing power from being applied to the chips. This would result in a condition where no testing could be performed on the chip until it was laser programmed. Because of yield loss, considerable wasted effort could be expended programming nonfunctional chips.

To allow power to be applied to the chip before laser programming, all of the vertical V_{DD} buses in the 64-by-8 memory slices have been isolated from the remainder of the V_{DD} bus structure. These vertical V_{DD} buses, which supply programmed V_{DD} to the memory elements, are electrically common and are brought to a common input test pad physically located between pads 21 and 22.

With the test pad floating or tied to ground, power can be applied to the ATL078 chip while all the programming links are still intact. This will allow all memory address locations to be accessed while leakage is monitored. Normally a logic "0" will be accessed from all memory elements, thus giving logic "1" output data on all 8 outputs. Should an address decoder, an NMOS memory element, or a data out multiplexer show a failure, the pull-up device(s) on the data output line(s) will pull up the internal data bus(es) and drive the tristate output(s) to a logic "0".

When the chip is to be operated normally, the test pad can be tied to V_{DD}.

21

H. SIMULATION

The entire critical data path of the ATL078 was simulated using a computer-aided simulation program. Simulation was heavily used in design and layout of the chip. Validation of transistor sizes and interconnect approaches was performed before they were incorporated into the chip.

The critical data path as shown in Fig. 7 was simulated in four pases at a 10 V operating voltage. With a 15 pF load on the output, the worst-case access time was found to be 117 ns. This increased to 126 ns with a 50 pf output load.

Table 2 shows a breakdown of the delays associated with each of the four parts of the critical access path. In each case, the slowest possible data path is considered. Figure 7 has the four simulated sections identified as AB, BC, CD and DE.

TABLE 2. SIMULATION OF WORST-CASE ACCESS PATH

Circuitry	Delays (ns)	
Input buffers to input of 1-of-64, including decoder gate RC.	22	
1-of-64 decoder driving NMOS memory matrix including row-select RC.	36	
NMOS memory matrix through multiplexers and sense amplifier.	42	
Intermediate amplifier and output tristate	17 (15 pF)	26 (50 pF)
Total	117 (15 pF)	126 (50 pF)

Section 4

CHIP STATISTICS

A summary of the ATL078 chip statistics is presented in Table 3. The ATL078 is a 4096-bit CMOS SOS ROM organized as 512 words by 8 bits. Each of the 8 outputs is implemented as tristate logic.

The technology used in laying out the ROM circuitry is the 7-mask SOS silicon gate technology. Standard design rules were used to implement the circuitry. These rules reflect a mature, produceable process. Provisions were made to program the ATL078 in either of two ways: either at the metal mask level (level 6) during processing, or by means of a directed laser beam after processing. Use of a laser beam to program the chip requires cutting metal links over the sapphire substate.

The total number of MOS transistors used to implement the ATL078 is 8782. The chip step-and-repeat dimensions are 252 by 257 mils, for a transistor density of 7.4 square mils per transistor. The chip fits in a 24-pin package such as the Metceram 80-0131. Only 23 of the 24 package pins are bonded to the chip. The chip has 24 pads, 23 bonded to the package pins and one used for pre-programing testing. This test pad is wired to V_{DD} in normal chip operation. The pinout used for this chip is directly compatible with the Intell 3604L-6 PROM and the 3304AL-6 ROM.

Computer simulations were made of the address and chip access times using the worst-case data path for each. The simulations were made at 25°C and at an operating voltage of 10 V. At a 15 pF external load, the chip-select access time is 35 ns, increasing 0.25 ns per pF as the load increases. The address access time was simulated as 117 ns at a 15 pF load. This delay also increases 0.25 ns per pF with addition external loading so that at 50-pf load the delay is 126 ns. The cycle time is identical to the worst-case address access time since the chip design uses only static logic.

Static power dissipation of the ATL078 is the sum of the standard CMOS leakage (across Off transistors) and the leakage contributed by the PMOS pull-up devices on the 8 data output lines. At 10 V, the PMOS pull-up leakage can vary from 0 to 24 mW. The normal CMOS Off transistor leakage should be 1 to 5 mW so that the total typical static leakage should be approximately 14 mW. A worst-case data condition could require 29 mW of static power, while a best-case dissipation could be as low as 1 mW.

TABLE 3. ATL078 CHIP STATISTICS

Parameter	Value or Characteristic	
Number of bits	4096	
Organization	512 x 8	
Outputs	Tristate	
Technology	CMOS SOS-Standard Design Rules	
Number of Process Masks	7	
Programming Options	Metal Mask or Laser	
Chip Size	252 x 257 mils	
Number of Transistors	8782	
Density	7. 4 sq. mils/transistor	
Number of I/O Leads	23 (uses 24-pin package)	
Package	Metceram 80-0131	
Pin Compatibility	Intel 3604L-6, 3304AL-6	
Address Access Time-WC* 10 Volt, 25°C	117 ns at 15 pF load 126 ns at 50 pF load	
Chip Select Access Time-WC* 10 Volt, 25°C	35 ns at 15 pF load 44 ns at 50 pF load	
Power (10 volts)	Typical	Worst Case
Static	14 mW	29 mW
Dynamic-no load	25 mW/MHz	40 mW/MHz
Cycle Time	Same as Address Access	

*worst-case

The dynamic dissipation on-chip is a sum of the CV^2f losses. The principal contributors to on-chip capacitance are the gate capacitance and the interconnect overlap capacitance. Summing this capacitance produced a typical calculated dynamic dissipation of 25 mW per MHz at 10 V. This assumes that 5 of the 9 address lines change state on every address update. Statistically this represents a high typical value. If all 9 address lines changed on every address transistion (not a likely case), the dynamic dissipation would not exceed 40 mW per MHz. The dynamic dissipation was calculated at no output load since output loading is system dependent and only increases the dynamic dissipation of the selected chips in a system. With an external loading of 50 pF on each of the 8 outputs of the chip, a typical dynamic dissipation of 20 mW per MHz would be expected (4 outputs changing). With all 8 outputs changing with every address change, a worst-case dynamic dissipation of 40 mW per MHz could be attributed to external loading.

Section 5

CONCLUSIONS

The design of this chip fills a void left by available ROM chip types; that of a high speed low power ROM that interfaces directly to CMOS systems.

The results of this 4K ROM development have produced a chip whose configuration, pinout, programming options, speed, and power meet the design objectives.

This 512-by-8 chip, designated the ATL078, has a pinout compatible with the Intel 3604L-6 PROM or the 3304AL-6 ROM. Programming is done at the metal mask level or by directed laser beam after processing. The ATL078 operating at 10 V is speed compatible to bipolar ROM chips at the system level, having a fast chip select access of 44 ns. A worst-case address access takes 126 ns.

At the system level, the power savings of the ATL078 compared to bipolar ROMs is 10:1 or better. Typical ROM systems should see average chip dissipations of less than 50 mW per chip, while 500 mW is typical of bipolar systems.

Section 6

RECOMMENDATIONS

The objective of this program was the design of a 4096-bit CMOS-SOS ROM that could be field programmed by a laser technique. Such a design was completed and documented. The method of programming by the laser scribing of metal links was previously demonstrated on several devices. A practical technique for programming this ROM would involve the modification of the laser scribing or cutting system to include an x-y step-and-repeat capability, which could be programmed to automatically scribe the metal link in accordance with the desired bit pattern. The cost of such systems was estimated as in the $60,000 to $100,000 range. Such a cost would tend to minimize the numbers of such programmers and consequently would severely limit their availability and usability. Therefore, such an approach is not recommended.

The need for a high speed (less than 100 ns access), 4K-bit or higher CMOS-SOS ROM in the tens of milliwatts range is greater now than ever - especially one that is reasonably hardened to total dose and dose rate. The CMOS-SOS technology is now more than capable of providing this rate of access, and basic designs for such a ROM have already been generated.

Because of the general need for such a low-power, high-speed radiation-resistant ROM in various space and air-borne applications, and because both the designs and technology to produce such a device exist, the design of such a CMOS/SOS ROM chip is highly recommended.

Appendix

PROGRAMMING LINK LOCATIONS

In order to laser cut the programming links on the 4096 memory bits in the ATL078, an accurate indication of their location must be given. This appendix provides that information.

Figure A-1 gives the relative topological layout of all the word and bit locations on the chip. A proper chip orientation places the chip identifer (ATL078) in the upper right-hand corner.

An explanation will be given of the information in the lower left hand quadrant of Fig. A-1. The leftmost and rightmost rectangles are labeled 0→63. These are the 1-of-64 decoders and the 0→63 indicates the direction of advancing address. Between the two 1-of-64 decoders is a rectangle with a vertically dashed center line, representing 64 by 16 bits of memory. The vertical dashing separates the 64-by-16 bits into two 64-by-8 slices. The numbers 320→383 indicate the true addresses contained in this 64-word slice. Proceeding from the bottom and going upwards, addresses 320-383 are on the left and addresses 256-319 are on the right. The order of the output bits is shown in the topmost rectangle of this quadrant. There are 16 numbers in this rectangle, representing two groups of 8 bits each. For example 2-1-3-4 . . . represents output bits 0_2-0_1-0_3-0_4. The three other quadrants can be interpreted in similar fashion.

Each bit of memory has two metal links associated with it. One link must be severed to program the bit. The location of the program links is repeatable in an array of 2 bits by 2 bits. Figure A-2 shows the repeatable 2-by-2 array. The two links associated with each bit are labeled "1" and "0". To program a bit so that the chip output is a high (logic "1") voltage the link labeled "1" must be severed. To get a low-level chip output, link "0" must be severed. The two links of any one bit have a horizontal separation of 0.8 mil. The links are 0.4 mil wide. Proceeding horizontally across a row, the link polarities do not strictly alternate; the polarities are 1-0-0-1-1-0-0-1- . . . and can be seen to be repeatable in groups of 4 links (2 bits). Every 2 bits is repeatable in 5.0-mil spacings in X, and 3.1-mil spacings in Y.

The location of the bottom edge of the "1" link (on its center-line) for address 320 output bit 0_2 is X50.9 Y16.2. This references a datum 0 at the intersection of the centerlines of the scribe lines in the lower left of the chip. (Note: to sever link "1", cut 0.4 mil in Y; to sever link 0, move 0.8 mil in X and cut 0.4 mil in Y.)
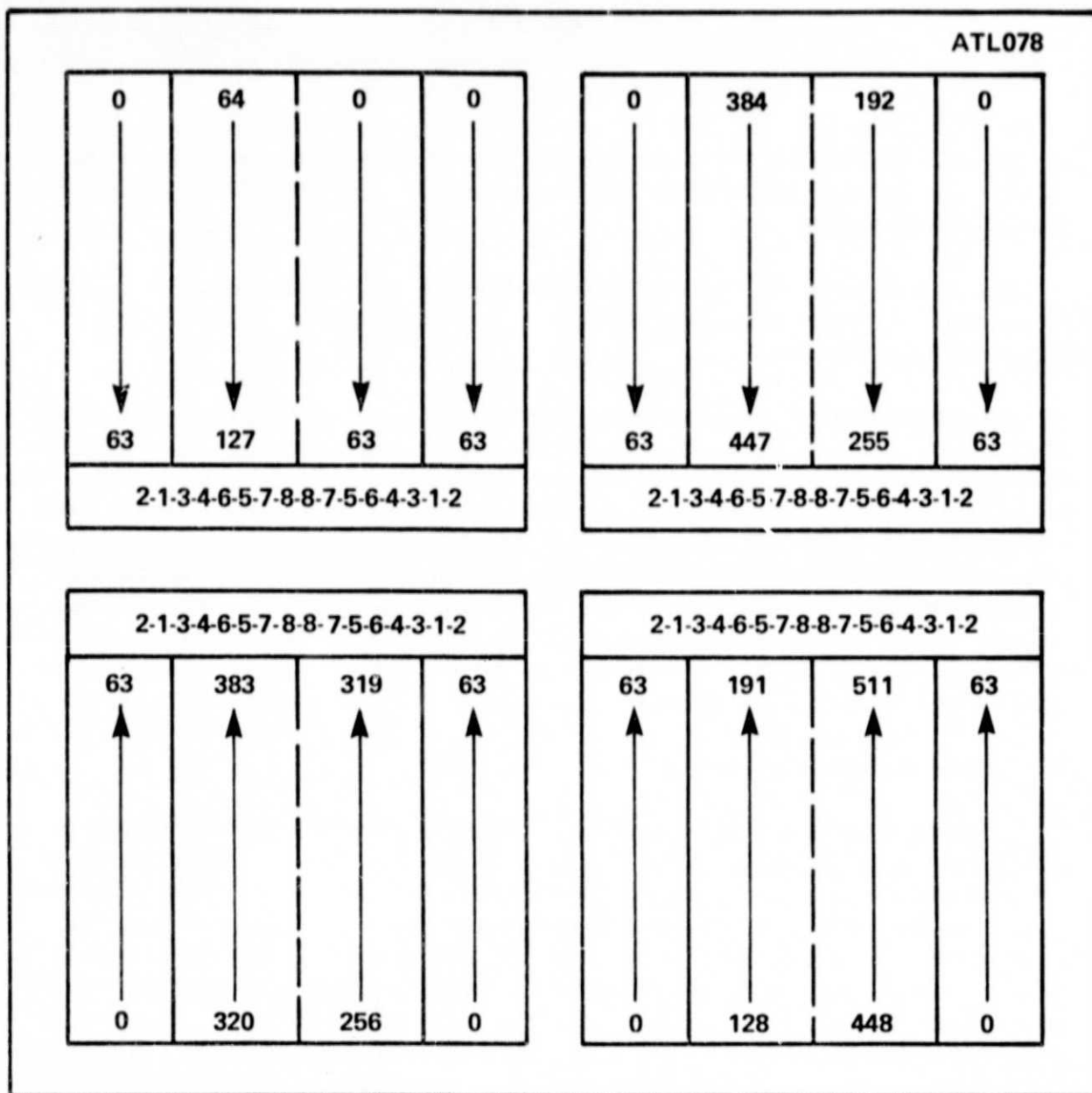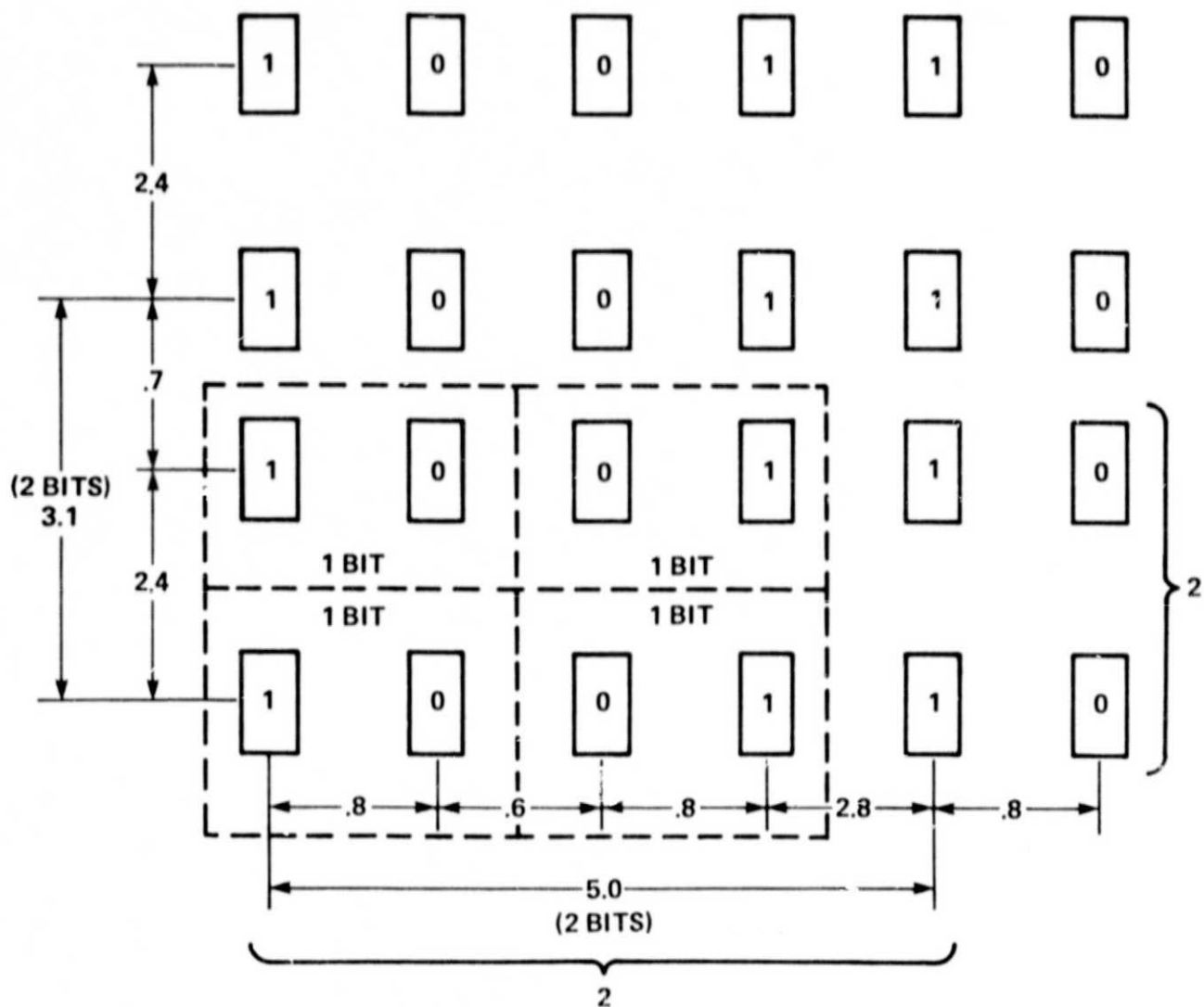
Fig. A-1. ATL078 word and bit locations.

Fig. A-2. Repeatable 2 x 2 array of program links.

The locations of the lower left "1" links in the other 3 quadrants are as follows:

Address	Bit	Location "1" Link
128	2	X162 Y16.2
127	2	X50.9 Y138.6
447	2	X162 Y138.6